

Multi-tensor Completion for Estimating Missing Values in Video Data

Chao Li^{*†}, Lili Guo^{*}, and Andrzej Cichocki[†] *Fellow, IEEE*

^{*}College of Information and Communication Engineering, Harbin Engineering University, China 150001

[†]Laboratory for Advanced Brain Signal Processing, RIKEN Brain Science Institute, Japan 351-0198

Email: {chao.li, a.cichocki}@riken.jp

Abstract—Many tensor-based data completion methods aim to solve image and video in-painting problems. But, all methods were only developed for a single dataset. In most of real applications, we can usually obtain more than one dataset to reflect one phenomenon, and all the datasets are mutually related in some sense. Thus one question raised whether such the relationship can improve the performance of data completion or not? In the paper, we proposed a novel and efficient method by exploiting the relationship among datasets for multi-video data completion. Numerical results show that the proposed method significantly improve the performance of video in-painting, particularly in the case of very high missing percentage.

I. INTRODUCTION

In computer vision and graphics, many methods were developed for image and video in-painting. Since not only RGB images but also videos can be considered as multi-mode arrays, more and more attention was paid on tensor-based algorithms [1] [2] [3]. However, it should be noted that the majority of methods cannot have a satisfactory performance when too much data was missed (e.g. the missing percentage of pixels is higher than 95%). It is because remaining observations have been not sufficient to predict missing values well. Fortunately, in many real scenarios, more than one device is usually used for observation, and all the datasets are mutually related. Thus it is straightforward to infer that we can achieve a better performance for image and video in-painting by exploiting the relationship among multi-datasets.

In this paper, we proposed a novel method for multi-tensor completion. We individually applied a low-rank approximation on every unfolding matrix along each mode of data, and folded estimations together as the prediction of missing values. More details will be discussed in Section 2. Additionally, related work will be introduced in Section 3, and Section 4 will provide results of numerical experiments.

II. OBJECTIVE AND ALGORITHM

Suppose that there are K tensors $\underline{\mathbf{X}}_k$, $k = 1, \dots, K$ with missing values which are indexed by binary tensor $\underline{\mathbf{W}}_k$, $k = 1, \dots, K$ (0-unobserved, 1-observed). In order to complete datasets, for each individual tensor $\underline{\mathbf{X}}_k$, the corresponding estimation $\underline{\mathbf{Y}}_k$ can be obtained by optimizing a traditional objective as [1] [3]

$$\begin{aligned} \min \quad & f(\underline{\mathbf{Y}}_k) = \|\underline{\mathbf{Y}}_k\|_* \\ \text{s.t.} \quad & P_{\underline{\mathbf{W}}_k}(\underline{\mathbf{Y}}_k) = P_{\underline{\mathbf{W}}_k}(\underline{\mathbf{X}}_k) \end{aligned} \quad (1)$$

where $\|\cdot\|_*$ denotes nuclear norm of a tensor, and operator $P_{\underline{\mathbf{W}}_k}(\cdot)$ denotes choosing observed elements of which the corresponding elements in $\underline{\mathbf{W}}_k$ are equal to 1. In multi-tensor case, and by the definition of the unclar norm of a tensor in [4], we have

$$\begin{aligned} \min \quad & f(\underline{\mathbf{Y}}_1, \dots, \underline{\mathbf{Y}}_K) = \sum_{k=1}^K \sum_{l=1}^{L_k} \alpha_{k,l} \|\mathbf{Y}_k^{(l)}\|_* \\ \text{s.t.} \quad & P_{\underline{\mathbf{W}}_k}(\underline{\mathbf{Y}}_k) = P_{\underline{\mathbf{W}}_k}(\underline{\mathbf{X}}_k), \quad k = 1, \dots, K \end{aligned} \quad (2)$$

where $\mathbf{Y}_k^{(l)}$ denotes unfolding tensor $\underline{\mathbf{Y}}_k$ along the l th mode, L_k , $k = 1, \dots, K$ is equal to the number of modes of $\underline{\mathbf{X}}_k$, and $\alpha_{k,l}$, $k = 1, \dots, K$, $l = 1, \dots, L_k$ are weights satisfying $\sum_{l=1}^{L_k} \alpha_{k,l} = 1$, $\forall k$. In order to search the optimal point of Eq.2 efficiently, we use a low-rank factorization with Frobenius norm regularization to approximate each unfolding of tensors. Specifically, we have an equivalent form of Eq.2 as

$$\begin{aligned} \min \quad & f(\{\underline{\mathbf{Y}}_k, \mathbf{U}_{k,l}, \mathbf{V}_{k,l} | \forall k, l\}) \\ & = \sum_{k=1}^K \sum_{l=1}^{L_k} \alpha_{k,l} \|\mathbf{Y}_k^{(l)} - \mathbf{U}_{k,l} \mathbf{V}_{k,l}^T\|_F^2 \\ & + \lambda \|\mathbf{U}_{k,l}\|_F^2 + \lambda \|\mathbf{V}_{k,l}\|_F^2 \end{aligned} \quad (3)$$

$$\text{s.t.} \quad P_{\underline{\mathbf{W}}_k}(\underline{\mathbf{Y}}_k) = P_{\underline{\mathbf{W}}_k}(\underline{\mathbf{X}}_k), \quad k = 1, \dots, K$$

where $\lambda > 0$ is tuning parameter, and $\mathbf{U}_{k,l}, \mathbf{V}_{k,l}$, $\forall k, l$ denotes factor matrices for $\mathbf{Y}_k^{(l)}$, $\forall k, l$. It is well known that Frobenius norm regularization on each factor matrices $\mathbf{U}_{k,l}, \mathbf{V}_{k,l}$, $\forall k, l$ results in a low-rank approximation of $\mathbf{Y}_k^{(l)}$, $\forall k, l$ [5]. Thus Eq.3 can be considered as a equivalent form of Eq.2.

For related multi-tensor, it is straightforward to infer that datasets would share information along some modes. Thus we suppose that there exists $k_1 \neq k_2$, or $l_1 \neq l_2$, such that $\mathbf{U}_{k_1, l_1} = \mathbf{U}_{k_2, l_2}$. It will be found that this assumption plays the key role to improve the performance for completion. To search the local optimal point of Eq.3, we can alternately update $\mathbf{U}_{k_0, l_0}, \mathbf{V}_{k_0, l_0}, \underline{\mathbf{Y}}_{k_0}$, $\forall k_0, l_0$ by

$$\mathbf{U}_{k_0, l_0} \leftarrow \left(\sum_{\substack{\forall k, l, \text{ s.t.} \\ \mathbf{U}_{k, l} = \mathbf{U}_{k_0, l_0}}} \alpha_{k, l} \mathbf{Y}_k^{(l)} \mathbf{V}_{k, l} \right) \cdot \left(\sum_{\substack{\forall k, l, \text{ s.t.} \\ \mathbf{U}_{k, l} = \mathbf{U}_{k_0, l_0}}} \alpha_{k, l} \mathbf{V}_{k, l}^T \mathbf{V}_{k, l} + \lambda \mathbf{I} \right)^{-1}, \forall k_0, l_0 \quad (4)$$

$$\mathbf{V}_{k_0, l_0} \leftarrow \left(\alpha_{k_0, l_0} \mathbf{Y}_{k_0}^{(l_0)T} \mathbf{U}_{k_0, l_0} \right) \cdot \left(\alpha_{k_0, l_0} \mathbf{U}_{k_0, l_0}^T \mathbf{U}_{k_0, l_0} + \lambda \mathbf{I} \right)^{-1}, \forall k_0, l_0 \quad (5)$$

$$\underline{\mathbf{Y}}_{k_0} \leftarrow \sum_{l=1}^{L_{k_0}} \alpha_{k_0, l} \text{ten}_l (\mathbf{U}_{k_0, l} \mathbf{V}_{k_0, l}^T) \quad \forall k_0 \quad (6)$$

$$P_{\underline{\mathbf{W}}_{k_0}} (\underline{\mathbf{Y}}_{k_0}) \leftarrow P_{\underline{\mathbf{W}}_{k_0}} (\underline{\mathbf{X}}_{k_0}) \quad \forall k_0 \quad (7)$$

where operator $\text{ten}_l(\cdot)$, $\forall l$ denotes transforming a matrix into a tensor which is opposite to the operator $(\cdot)^{(l)}$, $\forall l$, and \mathbf{I} denotes a identity matrix. In the algorithm, Eq.4-7 are alternately used to update $\mathbf{U}_{k, l}$, $\mathbf{V}_{k, l}$, $\underline{\mathbf{Y}}_k$, $\forall k, l$ until convergence. For initialization, choosing both $\mathbf{U}_{k, l}$, $\mathbf{V}_{k, l}$, $\forall k, l$ randomly and by Singular Value Decomposition (SVD) are recommended, and missing elements in $\underline{\mathbf{Y}}_k$, $\forall k$ can be initialized by zero. Compared to other traditional nuclear norm based methods, our method is non-convex. But it can stably provide a good performance in practice.

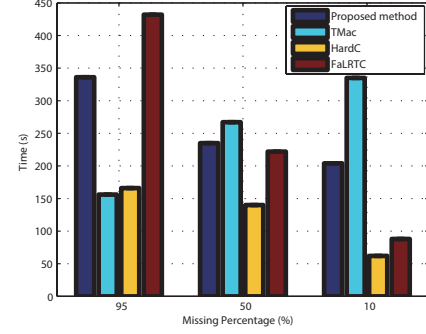
III. RELATED WORK

Many tensor based methods for completion have been developed [2] [1] [3] [6] [7], where [2] and [7] are based on tensor decomposition to predict missing values while [1] [3] [6] minimize nuclear norm of a tensor directly. But all of them just focus on a single tensor data. It is worthwhile to notice that TMac proposed in [6] used a similar optimization method to our method. However, we use the Frobenius norm regularization on factor matrices to control the rank of unfolding of tensors but TMac only justifies the number of column of latent matrices to control the nuclear norm.

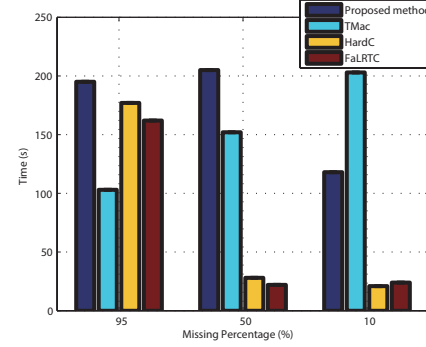
In many different field, multi-data analysis or group analysis was also deeply discussed [8] [9] [10]. But the majority of methods focus on common and distinctive component extraction. In this paper, we do not care latent features in each dataset. Thus we can reduce the complexity of algorithms compared to tensor decomposition based method like [2] [7]. A multi-matrix completion method Convex Collective Matrix Factorization (cCMF) was proposed in [10]. Compare to cCMF, The proposed method in this paper can be considered as an extension of cCMF into tensor problem.

IV. EXPERIMENTS

In this section, we use two experiments to evaluate the performance of the proposed method. In the first experiment,



(a) IXMAS



(b) RedSkirt

Fig. 1: Average running time of algorithms in two experiments

videos from IXMAS¹ were used to jointly predict missing values. IXMAS consists of videos in which 13 daily-live motions (e.g. sit down, check watch, kick, etc.) performed by 11 actors, actors can choose freely position and orientation, and each action was simultaneously recorded by 5 cameras. In the experiment, we choose 4 videos from two actors, they did the similar action (sit down), and there are two views for each actor. For each video, it can be modelled as a mode-3 tensor (Pixel \times Channel \times Time) $\underline{\mathbf{X}}_k$, $k = 1, 2, 3, 4$. Since two actors did similar action and each action was recorded by two cameras, videos recorded from the same camera can be supposed to have common pixel and channel information, and videos from the same actors share time-mode information. Under such an assumption, we implement the proposed method for video in-painting.

In the second experiment, we use multi-view videos (Height \times Width \times Channel \times Time) of human action (body move of a Asian lady) which were simultaneously recorded by 3 cameras². Since the three videos were recorded at the same time with different position, it is easy to infer that they share the information along the time mode. Further, we resize each video so that they lost common information from pixels, therefore tensor concatenation cannot be used for in-painting under this data. Such the case can be usually found

¹<http://4drepository.inrialpes.fr/public/viewgroup/6>

²<http://media.au.tsinghua.edu.cn/index.jsp>

TABLE I: RSE of prediction of multi-video in-painting in different missing percentage with 10 runs

Methods	95%		50%		10%	
	Data 1	Data 2	Data 1	Data 2	Data 1	Data 2
Proposed method	0.085	0.270	0.018	0.115	0.006	0.082
FaLRTC	0.116	0.403	0.016	0.129	0.006	0.078
HardC	0.118	0.559	0.023	0.133	0.020	0.079
TMac	0.145	0.331	0.033	0.280	0.020	0.271
CPWOPT	0.145	-	0.054	-	-	-

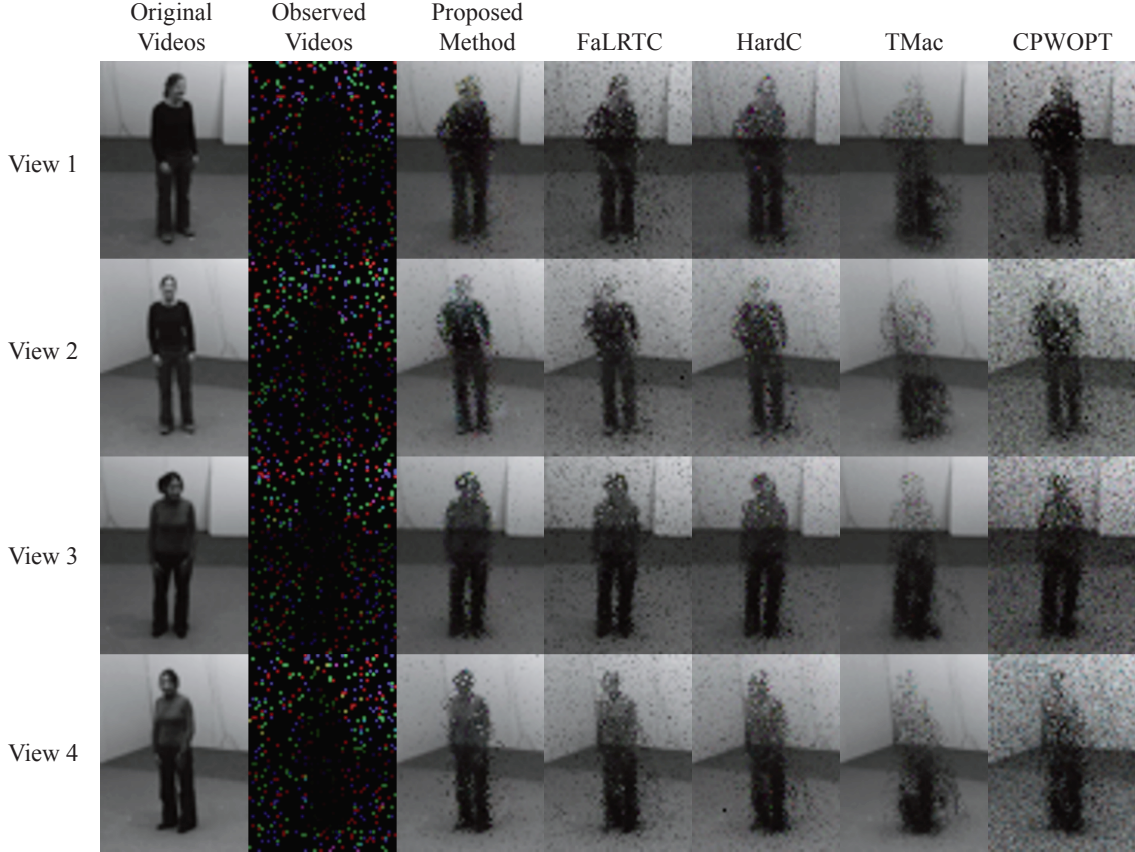


Fig. 2: Comparison of algorithms in the first frame of IXMAS

when different type of cameras were used in real application. Tab.I shows the Relative Square Error (RSE) for missing value prediction with different missing percentage, and Fig.1 shows comparison of the average running time from two experiments.

For comparison, FaLRTC [1], HardC [3], TMac [6], and CPWOPT [2] was individually implemented on every video. It is shown from Tab.I that the proposed method outperform other methods, particularly in the case of high missing percentage. It is because, in the case of high missing percentage, remaining information from observations for each individual video are too little for completion, but the proposed method can exploit relationship from all the datasets, and such the relationship provides more information to each dataset for missing value prediction. It is seen from Fig.1 that the running time of the proposed time is comparable to TMac, HardC, and FaLRTC,

while tensor-decomposition based algorithm CPWOPT which is not shown in Fig.1 are much slower than the other four methods. Fig.2 shows completion results from all the methods to estimate the first frame of IXMAS. Fig.2 also reflects that the proposed method can achieve a better performance for video in-painting.

V. CONCLUSION

In this paper, we developed a novel completion method for multi-tensor. Compared to traditional approaches, the proposed method can exploit common information shared among datasets. By numerical results, it is demonstrated that the shared information can improve the performance in multi-videos in-painting particularly in the case of high missing percentage.

ACKNOWLEDGMENT

This paper was supported by the China Scholarship Council.

REFERENCES

- [1] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 1, pp. 208–220, 2013.
- [2] E. Acar, D. M. Dunlavy, T. G. Kolda, and M. Mrup, "Scalable tensor factorizations for incomplete data," *Chemometrics and Intelligent Laboratory Systems*, vol. 106, no. 1, pp. 41–56, 2011.
- [3] M. Signoretto, Q. T. Dinh, L. De Lathauwer, and J. A. Suykens, "Learning with tensors: a framework based on convex optimization and spectral regularization," *Machine Learning*, vol. 94, no. 3, pp. 303–351, 2014.
- [4] M. Signoretto, L. De Lathauwer, and J. A. Suykens, "Nuclear norms for tensors and their use for convex multilinear estimation," *Submitted to Linear Algebra and Its Applications*, vol. 43, 2010.
- [5] N. Srebro, J. Rennie, and T. S. Jaakkola, "Maximum-margin matrix factorization," in *Advances in neural information processing systems*, 2004, pp. 1329–1336.
- [6] Y. Xu, R. Hao, W. Yin, and Z. Su, "Parallel matrix factorization for low-rank tensor completion," *arXiv preprint arXiv:1312.1254*, 2013.
- [7] Q. Zhao, L. Zhang, and A. Cichocki, "Bayesian cp factorization of incomplete tensors with automatic rank determination," *CoRR*, vol. abs/1401.6497, 2014.
- [8] G. Zhou, A. Cichocki, and S. Xie, "Common and individual features analysis: Beyond canonical correlation analysis," *arXiv preprint arXiv:1212.3913*, 2012.
- [9] A. Klami, G. Bouchard, and A. Tripathi, "Group-sparse embeddings in collective matrix factorization," *arXiv preprint arXiv:1312.5921*, 2013.
- [10] G. Bouchard, S. Guo, and D. Yin, "Convex collective matrix factorization," 2013.